General Movement Assessment from videos of computed 3D infant body models is equally effective compared to conventional RGB Video rating

A. Sebastian Schroeder, Nikolas Hesse, Raphael Weinberger, Uta Tacke, Lucia Gerstl, Anne Hilgendorff, Florian Heinen, Michael Arens, Linze Dijkstra, Sergi Pujades, et al.

General Movement Assessment from videos of computed 3D infant body models is equally effective compared to conventional RGB Video rating

Schroeder, A. Sebastian[1*] ; Hesse, Nikolas[2*,#] ; Weinberger, Raphael[1,5] ; Tacke, Uta[1,3] ; Gerstl, Lucia[1] ; Hilgendorff, Anne[1] ; Heinen, Florian[1] ; Arens, Michael[2] ; Dijkstra, Linze J.[6] Pujades Rocamora, Sergi[7] ; Black, Michael[7] ; Bodensteiner, Christoph[2] ; Hadders-Algra, Mijna[6]

[1] Ludwig Maximilian University of Munich (LMU), Hauner Children's Hospital, Department of Paediatric Neurology and Developmental Medicine, Munich, Germany
[2] Swiss Children's Rehab, University Children's Hospital Zurich, Affoltern am Albis, Switzerland
[3] University Children's Hospital, Basel, Switzerland
[4] Ludwig Maximilian University of Munich (LMU), Comprehensive Developmental Care (CDeC), Munich, Germany
[5] Ludwig Maximilian University of Munich (LMU), Department of Epidemiology, Institute of Social Paediatrics and Adolescent Medicine, Munich, Germany
[6] University of Groningen, University Medical Center Groningen, Department of Paediatrics, Groningen, Netherlands
[7] Max Planck Institute for Intelligent Systems, Tübingen, Germany

*Authors contributed equally

work was done while NH was with Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Ettlingen, Germany.

All authors have approved the final article

Declarations of interest: none

Corresponding author:
A. Sebastian Schroeder
Ludwig Maximilian University of Munich (LMU), Hauner Children's Hospital, Paediatric Neurology, Developmental Medicine, Lindwurmstr. 4, 80337 Munich
Phone: +49 89 4400 55110
Fax: +49 89 4400 55111
Email: Sebastian.Schroeder@med.lmu.de

General Movement Assessment from videos of computed 3D infant body models is equally effective compared to conventional RGB Video rating

Highlights (3-5 bullet points in an extra document): maximum 85 characters, including spaces, per bullet point

- Presentation of markerless 3D capture and modelling of general movements
- High agreement between conventional and model-based General Movement Assessment
- Prediction of CP of conventional and model-based GMA is equally good.

Abstract:

Background: General Movement Assessment (GMA) is a powerful tool to predict Cerebral Palsy (CP). Yet, GMA requires substantial training challenging its broad implementation in clinical routine. This inspired a world-wide quest for automated GMA.

Aim: To test whether a low-cost, marker-less system for three-dimensional motion capture from RGB depth sequences using a whole body infant model may serve as the basis for automated GMA.

Study design: Clinical case study at an academic neurodevelopmental outpatient clinic.

Subjects: Twenty-nine high risk infants were assessed at their clinical follow-up at 2-4 month corrected age (CA). Their neurodevelopmental outcome was assessed regularly up to 12-31 months CA.

Outcome measures: GMA according to Hadders-Algra by a masked GMA-expert of conventional and computed 3D body model ("SMIL motion") videos of the same GMs. Agreement between both GMAs was tested using dichotomous and graded scaling with Kappa and intraclass correlations, respectively. Sensitivity and specificity to predict CP at ≥12 months CA were assessed.

Results: Agreement of the two GMA ratings was moderate-good for GM-complexity (κ=0.58; ICC=0.874 [95%CI 0.730;0.941]) and substantial-good for fidgety movements (FMs; Kappa=0.78, ICC=0.926 [95%CI 0.843;0.965]). Five children were diagnosed with CP (four bilateral, one unilateral CP). The GMs of the child with unilateral CP were twice rated as mildly abnormal with FMs. GM-complexity and somewhat less FMs, of both conventional and SMIL motion videos predicted bilateral CP comparably to published literature.

Conclusions: Our computed infant 3D full body model is an attractive starting point for automated GMA in infants at risk of CP.

Keywords: General Movement Assessment (GMA), high risk infants, early diagnosis, fidgety, automated motion analysis, cerebral palsy, Kinect

Abbreviations:

BSID: Bayley Scales of Infant Development

CA: Corrected Age

CP: Cerebral Palsy

DA: Definitely Abnormal GMA rating

FMs: Fidgety Movements

GA: Gestational Age

GMA: General Movement Assessment

GMs: General Movements

MA: Mildly Abnormal GMA rating

NS: Normal Suboptimal GMA rating

NO: Normal Optimal GMA rating

SMIL: Skinned Multi-Infant Linear Model

SMPL: Skinned Multi-Person Linear Model

RGB-D: Red Green Blue-Depth

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Cerebral palsy (CP) describes the most common physical disability in children. It occurs in approximately 2/1.000 live births [1]. The average age at which the diagnosis is made and communicated with the families has been shown to be approximately 11 months. For less severely affected children diagnosis may even be made as late as 24 months of age [2]. Yet, early and accurate detection of infants at very high risk of CP is a prerequisite for timely initiation of early intervention, i.e. intervention starting before the age of 6 months [3]. Early intervention programs use caregiver coaching and address the infant's sensorimotor development, attention, self-regulation, early communication skills and parental mental well-being [4, 5].

To detect very high risk of CP, it is recommended to use a combination of standardized tools in conjunction with clinical history. Before 5 months corrected age (CA), the most predictive tools to detect risk of CP are term-age cerebral magnetic resonance imaging (MRI) and the video-based Qualitative Assessment of General Movements (GMA) between 2 and 5 months CA [6-8]). Term-age MRI may, however, not serve as a screening tool in clinical practice, leaving GMA as the most predictive tool in everyday practice.

General movements (GMs) are spontaneous movements, which involve all body parts [9, 10]. They emerge during early fetal life and disappear when goal-directed motor behaviour emerges around 4-5 months CA. The form of typical GMs changes as a result of developmental transformations of the nervous system. In the last phase, at 2 to 5 months CA, GMs have a 'fidgety' character. Fidgety GMs (FMs) occur irregularly all over the body and consist of a continuous stream of tiny elegant movements. During each phase, typical GMs are primarily characterized by complexity and variation (in short: complexity) [10]. GMs are considered as abnormal when their complexity is reduced. [9, 10]. At fidgety age – the age at which GMA best predicts developmental outcome [6, 11] – atypical GMs are also characterized by a reduction or disappearance of the age-specific FMs. Therefore Hadders-Algra suggested that GMA at 2-5 months CA (fidgety age) is performed with a two-step procedure: (1) grading movement complexity and (2) assessing FMs [11, 12].

Unfortunately, GMA requires a high level of expertise in the raters, which impedes the application in broad clinical practice [13-17]. Therefore, it is desirable to develop an easy to use, automated screening tool that can be applied in general paediatric practice as an alternative for the expert dependent video ratings of GMs. To be generally applicable, this diagnostic tool should be sensitive and specific and should use low-cost, commercially available equipment that does not influence the infants' motions.

A primary prerequisite for automatic analysis of motor development is to capture infant movements with high precision. Motions are represented as continuous time-series of the kinematics of limb and trunk movements, e.g., 2-dimensional (2D) or 3-dimensional (3D) body

joint positions or joint angles. In the subsequent analysis stage, motion features have to be extracted from the captured movements to serve as input to machine learning classifiers, which can be used to identify infants with abnormal motor activity or to predict the risk of CP.

The first computerized approach towards automated detection of CP used a marker-based Vicon system to track movements in 3D [18]. Relevant characteristics of the captured movements were encoded in motion features based on limb velocities and accelerations. A machine learning classifier was trained to predict the CP risk based on these features [18]. Up to date, a wide variety of sensors with different properties have been used for capturing infant movements, e.g., body-worn miniaturized movement sensors [19, 20] or magnetic tracking systems [21]. Video-based motion tracking has the advantage that it does not require the attachment of sensors [22, 23].

Marschik et al. [24] developed the 'iDN Fingerprint Model' that combines multiple biological signals of the young infant, including a 3D video of GMs with a Red-Green-Blue Depth (RGB-D) camera. The Italian group of Orlandi and colleagues proposed a system based on the combination of audio and video data, i.e., the infant's spontaneous vocalizations and its GMs [25]. However, currently both systems' psychometric properties have not been reported.

Purely computer-based 2D video analysis for the assessment of GMs was first published by Adde and colleagues [26]. They developed a General Movement Toolbox (GMT), which uses the displacement of pixels from one video frame to the next for quantitative analyses of GMs occurring at 2-5 months CA, i.e., the fidgety GMs. Visual representations from GMT resulted in recognisable patterns of the fidgety movements. They reported that the application of the automatic movement classification in their mixed study group of 82 high and low risk infants born at term or preterm would reduce the need for further referral by 70%. Their study showed that video recordings can be used for qualitative and quantitative analyses of GMs [26]. More recently, the same group suggested on the basis of a group of 150 high risk infants that the analysis could be used to reduce the rate of infants needing intensive developmental monitoring [27]. However, two things should be noted. First, the system misclassified some children, e.g., missing those who needed monitoring (10%) or referring those who did not require intensive follow-up (20%). Second, the authors had to exclude infants with a specific type of moderately abnormal GMs (abnormal FMs) from the analysis as the algorithm could not deal with these movements.

More recently, two other groups reported on computer-based analysis of 2D GM-videos. The group of Orlandi used an approach based on a skin model for segmentation and large displacement optical flow (LDOF) [28]. The kinematic features extracted from the model were used to classify fidgety GMs as typical or atypical. Their retrospective study based on 127 videos of preterm infants showed a specificity of 99%, but a sensitivity of only 44% in predicting

CP [28]. The group of Marchi and colleagues reported an approach based on automated pose estimation to capture key aspects of GMs [29]. They used OpenPose - an open source computer vision software - developed by Cao et al. [30] to analyse the 2D videos of 21 preselected infants at fidgety GM-age. They considered their approach as feasible for automatizing infant motion analysis. However, they indicated that pose tracking errors occurred in 14 out of 21 recorded videos, which hindered automatized analysis. They concluded, that their approach would improve significantly by the use of three dimensional (3D) instead of 2D camera technology [29].

To summarize, up to now no results are available on the application of 3D systems for automated assessment of movements in infancy that do not use sensors/markers that have to be attached to the infants' limbs, which may affect their behaviour.

Since 2015, we have developed a motion tracking system that estimates the full body pose of infants in 3D [31-34] (Kinematic Motion Analysis Tool; KineMAT). The system is based on a commercially available, low-cost RGB-depth sensor (Kinect 1.0, Microsoft, USA) and can be applied without attaching markers to the infant's body (see Figure 1) [31-33]. Recently, we developed the Skinned Multi-Infant Linear (SMIL) body model [34], which is based on the adult body model SMPL (Skinned Multi-Person Linear Model) [35]. The pose and shape of this virtual body can be adjusted with the model parameters. This allows us to use SMIL for capturing the 3D shape and 3D pose of infants from RGB-D sequences. We create a video of the resulting SMIL body reproducing the movements of the recorded infant, which we denote "SMIL motion video". Figure 2 illustrates the resemblance of the RGB and SMIL motion videos.

The aim of this study was to evaluate how reliably our developed 3D infant body model (SMIL motion video) represents true infant spontaneous movements to a human observer. To this end, we assessed the agreement on the classifications of a human expert of GM-quality of two videos of the same GMs of 29 consecutively recruited high risk infants at fidgety age. The data of both videos originated from the same camera recording. One video consisted of the conventional GM-video used in clinical practice, the other of the SMIL motion video (Fig. 1 and Fig. 2). Both videos were assessed according to Hadders-Algra, i.e., assessing GM-complexity (in the present study with three degrees of fine-grading) and assessing FMs with two degrees of grading [12, 36, 37].

Material and Methods

Participants

Participants were high risk infants who received standardized follow up at the Centre of Developmental Care of the integrated Social Paediatric Centre at the University Hospital in Munich, Germany. High risk infants were infants meeting at least one of the following criteria: born ≤ 32 weeks of gestation (GA), birth weight below 1500 grams, or other perinatal complications such as term asphyxia or early onset meningitis. Exclusion criteria were additional, potentially confounding diagnoses (e.g. genetic disorders, brachial plexus palsy, neuromuscular disorders). All infants meeting the inclusion criteria were examined by the first author between November 2016 and October 2017. The video recording was performed prior to the general developmental neurological examination. The infants' prenatal, perinatal and neonatal history was documented on standardized charts. After fidgety GM-age, the infants generally were assessed at a three to six months interval depending on the clinical need. Each follow-up visit consisted of a detailed neuropaediatric clinical examination. In children ≥12 months CA the diagnosis of CP was based on clinical findings in line with international recommendations as published by Boychuk et al. [38]. At 24 months CA, children were invited to participate in a structured neurodevelopmental assessment using the Bayley Scales of Infant and Toddler Development (BSITD, Bayley III Scales, full version).

The local ethics committee gave clearance for this study prior to the inclusion of the first patient (Ethikkommission LMU, Project Nr. 454-16). Parents gave written informed consent prior to study entry. No family withdrew consent during the study period.

Methods

The infant lay in supine position and an RGB-D camera (Microsoft Kinect V1, USA) was positioned directly above the infant at a distance of 100 cm. This camera allows simultaneous recordings of regular RGB videos (typically used for clinical GMA ratings) and depth videos, which provide the 3D distance to the camera for each pixel of the image. The infant's GMs were recorded for 3 minutes in a peaceful, quiet environment during active wakefulness.

To capture the infant's motion, we use SMIL (Skinned Multi-Infant Linear body model; see introduction for additional details on the development of the model [34]), which models the 3D body surface together with an underlying skeleton. The pose and shape of this virtual body can be adjusted with the model parameters. This allows the "registration" of the model to RGB-D data. Registration denotes the process of (automatically) adjusting model parameters so that the model best describes the input data, i.e. the model takes the same shape and pose as the recorded infant. We create a video of the SMIL registration results, i.e., the SMIL body reproducing the recorded infant movements, which we denote as "SMIL motion video" (see Fig. 1 and Fig. 2). SMIL allows access to 3D positions of all points of the body and the skeleton,

as well as body joint angles in all degrees of freedom across time. The SMIL Model achieves a metric accuracy, i.e., the average distance of RGB-D scan points to the model surface, of 2.51 mm (SD 0.21 mm) and represents the original (correct) infant body pose in 98.8% of the time [34]. The model is adapted to the specifics of infantile body proportions. Our registration method can handle fast movements and self-occlusions, which are a common problem when capturing motions of freely moving humans with only one camera. Further details on the registration of the SMIL Model to RGB-D sequences can be found in previous publication [34].

The conventional video and the SMIL motion video of each infant were used for blinded GMA rating according to Hadders-Algra. For GMA at fidgety age this means a two-step procedure. First, the degree of GM-complexity is assessed – this is the parameter that can be assessed at any GM-age [10]. It supplies information on the integrity of the subcortical-cortical networks [11]. We graded GM-complexity in three ways: (a) by using a classification into four categories: normal-optimal movements (NO; abundant complexity and variation), normal-suboptimal movements (NS; sufficient complexity and variation), mildly abnormal movements (MA; insufficient variation and complexity; reflecting normal but non-optimal brain function) and definitely abnormal movements (DA; very limited complexity and variation including cramped synchronized movements; reflecting brain dysfunction) [10] (b) by applying a fine-graded 10-point Likert-scale, in which score 1 denotes a virtual absence of GM-complexity and 10 a very abundant GM-complexity [37]. Scores of NO movements ranged from 10 to 8, those of NS movements from 7 to 6, those of MA movements from 5 to 4 and those of DA movements from 3 to 1. The Likert-scale fine grading has been applied before; it helps to describe movement quality but it does not assist improvement of prediction of developmental outcome compared to that based on the four classes [37]. (c) GMA ratings were dichotomized into the clinically relevant DA category (1-3 on the Likert scale; high predictive value for CP) versus a non-DA category (MA, NS, NO, i.e. 4-10 on the Likert scale) as previously described by Hadders-Algra et al. [39]. Second, the degree of FMs is assessed: continuously present, intermittently present, sporadic and absent [40]. Continuously and intermittently present FMs were considered as typical FMs. In the calculation of predictive values FMs were categorized as absent or present, i.e., sporadic or typical. [12].

The senior author assessed all SMIL motion videos in random order (both in terms of GM-complexity and in terms of presence of FMs), followed by the assessment of all RGB videos in random order. The senior author is a GMA expert with more than 25 years of experience in clinical and scientific GMA rating (MH-A). She was masked for the infant's medical history and outcome.

Statistical analysis

Descriptive statistics were used to describe the infants' clinical information on single subject level (gestational age at birth, APGAR, umbilical cord pH and base excess, age at GMA recording, neonatal morbidity diagnoses, brain ultrasound findings, results of neurodevelopmental outcome at ≥12 months, GMA scores).

For the assessment of both videos by the same GMA rater, statistical tests representing test-retest and intra-rater reliability measures were used. The Intraclass Correlation Coefficient (ICC; two way mixed, absolute agreement) was used to evaluate the degree of agreement between the two GM-ratings of GM-complexity using the 1-10 Likert scale and the FM-ratings. Reliability is considered high when ICC is 0.70 or higher. The ICC values are presented with 95% confidence intervals (CI). In addition, we used Bland-Altman plots to report the agreement between the GM-complexity on the basis of both videos.

Kappa statistics were used to describe the agreement of DA and Not-DA as well as for the classification of absent and present FMs of both types of video. Kappa-values of 0.41-0.60 are considered a "moderate agreement", 0.61-0.80 a "substantial agreement" and values of 0.81-1.00 as "almost perfect agreement". Two-by-two tables were used to describe the consistency of GMA for the conventional videos compared to those based on the SMIL motion video with respect to neurodevelopmental long-term outcome (CP present or absent). In addition, specificity, sensitivity, negative predictive value, and positive predictive value were calculated. Statistics were performed with IBM SPSS, version 25.0.0.1 (IBM Corporation USA). ROC curves were used to graphically illustrate the relationship between sensitivity and the occurrence of false positive values for DA versus non-DA GM-complexity ratings of both types of video.

Results

Twenty-nine infants met the inclusion criteria. Thus, 2x29 sets of videos were assessed. The infants' mean corrected age at recording was 14.8 weeks (95%CI 14.1;15.5). More than half (18/29, 62%) of the infants was born before 32 weeks of gestation. All infants born at higher gestational ages were highrisk due to perinatal or postnatal complications (e.g. term asphyxia or meningitis) requiring standardised neurodevelopmental follow up examinations. At neurodevelopmental follow-up between 12-31 months CA 10 children (34%) showed a typical development; 14 had a global developmental delay (48%) and five (17%) had an ascertained diagnosis of CP (four a bilateral spastic CP, one a unilateral spastic CP). Detailed clinical characteristics of the infants are provided in **Table 1 and Supplementary Table S1**.

GMA rating: conventional video versus SMIL motion video

On the basis of GM-complexity assessments six infants were rated as DA, both in the conventional videos and in the SMIL motion video rating, 18 (13 SMIL motion video) as MA, and 6 (10) as NS. In both video ratings no infant was rated as NO (**Table 1**).

Using the dichotomy DA versus non-DA GM-complexity, 25/29 (86%) of both ratings were identical**.** The corresponding inter-rating agreement measured with the Cohen's Kappa was 0.58 (p = 0.002), meaning "moderate" agreement. For FMs present versus absent Cohen's Kappa was 0.78 (p = 0.001), meaning "substantial" agreement.

The GM-complexity ratings of both video methods using the 1-10 Likert scale were identical for 16/29 (55%) infants. An additional 10/29 (34%) ratings showed only a +/-1 difference between both video-methods. The ratings of three infants (10%) differed more than one point. In two infants the ratings differed by two points (#6: coventional GMA rating of 5, corresponding to MA GM-complexity, and a SMIL motion video GMA rating of 3, corresponding to DA GM-complexity; *#15* conventional GMA rating of 3 and a SMIL motion video GMA rating of 5). In the third infant the ratings differed by three points (#3, conventional GM-complexity 5 versus SMIL motion video GM-complexity 2) (**Table 2, Bland-Altman-Plot see Supplementary Figure S1**). The mean Intraclass Correlation Coefficient (ICC) of both GM-complexity ratings was excellent 0.874 (95% CI 0.730;0.941).

In the FM rating of the conventional video FM's were absent in three infants (#1, #2, #6) of which two infants had DA GM-complexity and one MA GM-complexity. All three were later diagnosed with CP. In the SMIL motion video also three infants were rated as absent FM (#1, #6, #8) of which again two infants were rated having DA GM-complexity and one as having MA GM-complexity. Two of these were diagnosed with CP and one with global developmental delay at follow up. ICC of FM ratings was 0.926 (95% CI 0.843;0.965).
One infant who was diagnosed with bilateral CP (#5) and the infant with unilateral CP (#20) had presented either with sporadic or typical FMs in both videos (**Table 1**).

Prediction of developmental outcome

Table 2 summarizes the predictive values of the various GMA ratings including positive and negative predictive values. Sensitivity and specificity of DA vs non-DA rating of GM-complexity on the basis of the conventional videos to predict CP were 0.600 and 0.875, respectively; those of the corresponding SMIL motion videos 0.800 and 0.917, respectively. The difference between the two is illustrated by ROC curves (Supplementary Figure S2). Sensitivity and specificity of absent FMs of the conventional video were 0.600 and 1.000, those of the SMIL motion videos 0.400 and 1.000, respectively. Finally, the combination of GM-complexity and FM ratings resulted in a sensitivity and specificity of both types of video of 0.400 and 1.000,

respectively. Child #20, later diagnosed with unilateral CP, was consistently missed: she showed MA GMs with FMs.

Discussion

Our clinical case-study indicated that GMA based on computer generated virtual 3D infant body models (SMIL motion videos) closely corresponds to the established gold standard based on conventional RGB videos. The SMIL motion videos were able to catch movement complexity and FMs. In addition, our data suggest that GMA ratings based on SMIL motion videos result in a similar prediction of CP as GMA based on conventional videos.

Comparison of GMA based on conventional versus SMIL motion videos

Several studies have reported excellent inter-observer agreement of skilled observers for conventional GMA [10, 41, 42]. Our comparison of the 29 simultaneously recorded video pairs could be considered a similar type of test-retest analysis. Corresponding to literature data, we found substantial inter-observer reliability (dichotomous FMs and GM-complexity Cohen's Kappa 0.78 and 0.58 respectively, finer graded ratings ICC 0.926 and 0.874 respectively).

The predictive values of both ratings of GM-complexity were comparable to those described in the literature; those of FMs were somewhat lower than the values reported in the literature, where summary estimates of sensitivity and specificity of GMA based on FM assessment have been reported to be 98% [95% CI: 74;100] and 91% [95% CI: 83;93%] respectively [6]. Interestingly, the predictive values based on the SMIL motion video were not inferior to those based on the conventional videos. While GM-complexity rating seemed to be superior using the SMIL motion video (Sensitivity 0.8 versus 0.6; PPV 0.667 versus 0.5, NPV 0.957 versus 0.913), the conventional video seemed to be slightly superior regarding the assessment of FMs (Sensitivity 0.6 versus 0.4; NPV 0.923 versus 0.889). It is conceivable that the SMIL motion video is superior to conventional GMA rating of GM-complexity, as the SMIL motion video forces the observer to focus on the Gestalt perception of GM-quality, whereas in the conventional video the assessor may be distracted and/or biased by additional visible details, such as naso-gastric tube, oxygen supply, monitoring cables or details of facial expression. Regarding the assessment of FMs, however, the SMIL motion video at times contains slight pose inaccuracies due to noisy input data, which may be misinterpreted as very small amplitude FMs. This could explain that it was occasionally hard to distinguish sporadic and absent FMs on the basis of the SMIL videos due to insufficient data quality. This underpins the conclusion, that assessing both parameters (GM-complexity and FMs) is relevant for an optimal clinical prediction.

To understand the differences in the GMA ratings based on the conventional and SMIL motion videos of infants #2, #3, #6, and #15 MH-A reviewed these videos. She noticed the following for the infants (#3 and #6) whose GM-complexity was rated better on the conventional video (MA) than on the SMIL video (DA): Infant #3 (developmental delay at follow-up) showed especially some complex movements in the distal joints (with FMs), apparently these movements made a less prominent impression on the SMIL video. Infant # 6 who was later diagnosed with bilateral CP had shown on the conventional video a discrepancy in movement complexity between the upper and the lower part of the body, with the upper being better than the lower. This discrepancy was less striking on the SMIL video. Both videos were rated as absent FMs. Infant #15 (typical outcome at follow-up) had been assessed at 9 weeks CA. GM-complexity of the conventional video was rated as DA, due to prominent stereotyped movement sequences in the proximal joints; the distal joints showed some complexity. The latter was more prominently visible during the SMIL video, which was therefore rated as MA. Both videos were rated as sporadic FMs. It has been shown by others, that movement quality at 9 weeks CA may improve to better qualities during the following weeks [43]. Presumably this happened also in this infant. Finally, the GMs of infant #2 (bilateral CP) were rated on both videos as DA. However, the FMs were scored as absent on the conventional video, but as sporadic on the SMIL video – the few FMs had been observed at the wrists only. The latter may also be due to the slight pose inaccuracies due to noisy input data of the SMIL Model suggesting the presence of sporadic FM. The differences in scores between conventional and SMIL videos could be due to such technical matters, but they also underline that GMA based on Gestalt perception may remain difficult even after years of experience; it stresses the need of automatic movement analysis following strict algorithms.

GMA was especially adequate in the prediction of bilateral CP. Yet, it missed the child who was diagnosed with unilateral CP. It has been reported before that GMA does not detect all children with unilateral CP [44-46]. The differential ability of GMA to predict bilateral and unilateral CP corresponds to the putative neural substrate of the quality of GMs. GM-quality is considered to reflect the integrity of extensive neural networks involving not only cortical areas, but also their connectivity with subcortical relay stations [7]. Indeed, evidence is accumulating that DA GMs are especially related to damage of the periventricular white matter [7, 47]. In part of the children with unilateral CP the lesion is restricted to the cortical grey or does only affect part of the cross-roads running through the periventricular white matter (as in infant #20) and does not massively disrupt white matter connectivity, therewith not producing DA GMs. The detection of children with unilateral CP is also challenging due to the finding that fidgety movements are frequently present.

Two children showed DA GMs (#4 and #7) but were not diagnosed with CP. They both presented with "global developmental delay" at follow-up. Recent data indicate that the limitations in the infant's early motor repertoire do not only predict an increased risk of CP, but also an increased risk of cognitive impairment irrespective of the GMA method used for classification [48-50].

Strengths and limitations:

The strength of the current study is that it is the first that compares GMA based on videos of computed 3D infant body models to GMA rating of conventional videos. Another strength is that the videos were rated by one masked and highly experienced GMA assessor using multiple grading strategies including explicit assessment of FMs.

The study also has limitations: the sample size was small and consisted of high risk infants only. However, this convenience sample might be considered as representative of many infant follow-up clinics and represented infants with normal, MA and DA GMs. It may also be considered that the duration of follow-up was limited, which may mean that some children with a developmental diagnosis may have been missed [37].

Current status and future perspective of GMA based on SMIL modelling

In contrast to previous approaches that capture a limited number of body joint positions based on 2D-video [27-29, 51, 52] or relying on marker based sensors attached to the baby [18, 19, 53, 54], our method allows the extraction of angles of all body joints (3 degrees of freedom per joint), 3D positions of body joints and of arbitrary points on the body surface without the need of potentially distracting markers. Our method is robust to self-occlusions (e.g. hands and feet occluding large parts of the body) and fast motion. It has been proven to display the true whole body movements of the recorded infant in 98.8% of recorded video sequences, and to have a metric accuracy, i.e. the average distance of RGB-D scan points to the model surface, of 2.51 mm (SD 0.21 mm) [34].

In order to further improve the SMIL tracking, we plan to integrate means to detect and fix the rare occurrence of failure cases, as recently published by Aristidou et al. [55]. In addition, we aim to apply the SMIL motion video in a larger patient population, not only to improve the detection of predetermined motion parameters (e.g. fidgety general movements, centroid of motion, cramped synchronized activity, or a reduction of segmental movements), but also to use machine learning algorithms to quantify the motion parameters which are assumed to drive the "Gestalt perception".

Concluding remarks

Our case study of 29 high risk infants demonstrated that the amount of motion details captured by the SMIL motion video - based on a low-cost Kinect recording and the KineMAT tool - enables accurate GMA at fidgety age. This implies that the SMIL motion video adequately catches the movement characteristics needed for GMA of infants with movements ranging from a normal to a definitely abnormal quality, turning it into an attractive tool for automatic GMA.

References

1. Himmelmann K, Uvebrant P. The panorama of cerebral palsy in Sweden part XII shows that patterns changed in the birth years 2007-2010. Acta Paediatr. 2018;107(3):462-8.
2. Granild-Jensen JB, Rackauskaite G, Flachs EM, Uldall P. Predictors for early diagnosis of cerebral palsy from national registry data. Dev Med Child Neurol. 2015;57(10):931-5.
3. Novak I, Morgan C, Adde L, Blackman J, Boyd RN, Brunstrom-Hernandez J, et al. Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. JAMA Pediatr. 2017;171(9):897-907.
4. Akhbari Ziegler S, Dirks T, Hadders-Algra M. Coaching in early physical therapy intervention: the COPCA program as an example of translation of theory into practice. Disabil Rehabil. 2019;41(15):1846-54.
5. Hutchon B, Gibbs D, Harniess P, Jary S, Crossley SL, Moffat JV, et al. Early intervention programmes for infants at high risk of atypical neurodevelopmental outcome. Dev Med Child Neurol. 2019;61(12):1362-7.
6. Bosanquet M, Copeland L, Ware R, Boyd R. A systematic review of tests to predict cerebral palsy in young children. Dev Med Child Neurol. 2013;55(5):418-26.
7. Hadders-Algra M. Early human brain development: Starring the subplate. Neurosci Biobehav Rev. 2018;92:276-90.
8. Kwong AKL, Fitzgerald TL, Doyle LW, Cheong JLY, Spittle AJ. Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review. Dev Med Child Neurol. 2018;60(5):480-9.
9. Prechtl HF. Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction. Early Hum Dev. 1990;23(3):151-8.
10. Hadders-Algra M, Mavinkurve-Groothuis AM, Groen SE, Stremmelaar EF, Martijn A, Butcher PR. Quality of general movements and the development of minor neurological dysfunction at toddler and school age. Clin Rehabil. 2004;18(3):287-99.
11. Hadders-Algra M. Neural substrate and clinical significance of general movements: an update. Dev Med Child Neurol. 2018;60(1):39-46.
12. Hamer EG, Bos AF, Hadders-Algra M. Assessment of specific characteristics of abnormal general movements: does it enhance the prediction of cerebral palsy? Dev Med Child Neurol. 2011;53(8):751-6.
13. Adde L, Rygg M, Lossius K, Oberg GK, Stoen R. General movement assessment: predicting cerebral palsy in clinical practise. Early Hum Dev. 2007;83(1):13-8.
14. Ricci E, Einspieler C, Craig AK. Feasibility of Using the General Movements Assessment of Infants in the United States. Phys Occup Ther Pediatr. 2018;38(3):269-79.
15. Brown AK, Greisen G, Haugsted U, Jonsbo F. Formal training in general movement assessment is required to effectively evaluate infants with perinatal asphyxia in outpatient settings. Acta Paediatr. 2016;105(9):1056-60.
16. Bernhardt I, Marbacher M, Hilfiker R, Radlinger L. Inter- and intra-observer agreement of Prechtl's method on the qualitative assessment of general movements in preterm, term and young infants. Early Hum Dev. 2011;87(9):633-9.
17. Datta AN, Furrer MA, Bernhardt I, Huppi PS, Borradori-Tolsa C, Bucher HU, et al. Fidgety movements in infants born very preterm: predictive value for cerebral palsy in a clinical multicentre setting. Dev Med Child Neurol. 2017;59(6):618-24.
18. Meinecke L, Breitbach-Faller N, Bartz C, Damen R, Rau G, Disselhorst-Klug C. Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. Hum Mov Sci. 2006;25(2):125-44.
19. Heinze F, Hesels K, Breitbach-Faller N, Schmitz-Rode T, Disselhorst-Klug C. Movement analysis by accelerometry of newborns and infants for the early detection of movement disorders due to infantile cerebral palsy. Med Biol Eng Comput. 2010;48(8):765-72.

20. Gima H, Shimatani K, Nakano H, Watanabe H, Taga G. Evaluation of Fidgety Movements of Infants Based on Gestalt Perception Reflects Differences in Limb Movement Trajectory Curvature. Phys Ther. 2019;99(6):701-10.
21. Philippi H, Karch D, Kang KS, Wochner K, Pietz J, Dickhaus H, et al. Computer-based analysis of general movements reveals stereotypies predicting cerebral palsy. Dev Med Child Neurol. 2014;56(10):960-7.
22. Rahmati H, Martens H, Aamo OM, Stavdahl O, Stoen R, Adde L. Frequency-based features for early cerebral palsy prediction. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:5187-90.
23. Stahl A, Schellewald C, Stavdahl O, Aamo OM, Adde L, Kirkerod H. An optical flow-based method to predict infantile cerebral palsy. IEEE Trans Neural Syst Rehabil Eng. 2012;20(4):605-14.
24. Marschik PB, Pokorny FB, Peharz R, Zhang D, O'Muircheartaigh J, Roeyers H, et al. A Novel Way to Measure and Predict Development: A Heuristic Approach to Facilitate the Early Detection of Neurodevelopmental Disorders. Curr Neurol Neurosci Rep. 2017;17(5):43.
25. Orlandi S, Guzzetta A, Bandini A, Belmonti V, Barbagallo SC, Tealdi C, et al. AVIM—A contactless system for infant data acquisition and analysis: Software architecture and first results. Biomedical Signal Processing and Control. 2015;20:85-99.
26. Adde L, Helbostad JL, Jensenius AR, Taraldsen G, Stoen R. Using computer-based video analysis in the study of fidgety movements. Early Hum Dev. 2009;85(9):541-7.
27. Stoen R, Songstad NT, Silberg IE, Fjortoft T, Jensenius AR, Adde L. Computer-based video analysis identifies infants with absence of fidgety movements. Pediatr Res. 2017;82(4):665-70.
28. Orlandi S, Raghuram K, Smith CR, Mansueto D, Church P, Shah V, et al. Detection of Atypical and Typical Infant Movements using Computer-based Video Analysis. Conf Proc IEEE Eng Med Biol Soc. 2018; 3598-3601.
29. Marchi V, Hakala A, Knight A, D'Acunto F, Scattoni ML, Guzzetta A, et al. Automated pose estimation captures key aspects of General Movements at eight to 17 weeks from conventional videos. Acta Paediatr. 2019;108(10): 1817-1824.
30. Cao Z, Simon T, Wei SE, Sheikh Y. Realtime multi-person 2D pose estimation using part affnity fields. CVPR, IEEE. 2017:1302-10.
31. Hesse N, Stachowiak G, Breuer T, Arens M. Estimating Body Pose of Infants in Depth Images using Random Ferns. IEEE International Conference on Computer Vision Workshops (ICCVW). 2015:35–43.
32. Hesse N, Schroeder A, Mueller-Felber W, Bodensteiner C, Arens M, Hoffmann U. Markerless motion analysis for early detection of infantile movement disorders. EMBEC & NBC 2017, Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC): Tampere, Finland, June 2017. DOI: 10.1007/978-981-10-5122-7_50.
33. Hesse N, Schroeder A, Mueller-Felber W, Bodensteiner C, M. A, Hofmann U. Body Pose Estimation in Depth Images for Infant Motion Analysis. EMBC 2017, 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society : July 11 to 15, 2017, Jeju Island, Korea. DOI: 10.1109/EMBC.2017.8037221.
34. Hesse N, Pujades S, Black M, Arens M, Hofmann U, Schroeder S. Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences. IEEE Trans Pattern Anal Mach Intell. 2019.DOI: 10.1109/TPAMI.2019.2917908.
35. Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ. Smpl: A skinned multi-person linear model. ACM Transactions on Graphics (TOG). 2015;34(6):248:1-:16.
36. Hadders-Algra M. Evaluation of motor function in young infants by means of the assessment of general movements: a review. Pediatr Phys Ther. 2001;13(1):27-36.
37. Groen SE, de Blecourt AC, Postema K, Hadders-Algra M. General movements in early infancy predict neuromotor development at 9 to 12 years of age. Dev Med Child Neurol. 2005;47(11):731-8.

38. Boychuck Z, Andersen J, Bussieres A, Fehlings D, Kirton A, Li P, et al. International expert recommendations of clinical features to prompt referral for diagnostic assessment of cerebral palsy. Dev Med Child Neurol. 2020; 62(1): 89-96.
39. Hadders-Algra M. Putative neural substrate of normal and abnormal general movements. Neurosci Biobehav Rev. 2007;31(8):1181-90.
40. Einspieler C, Prechtl HF, Bos A, Ferrari F, Cioni G. Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants. First ed: Mac Keith Press; 2008 September 2008.
41. Hadders-Algra M, Groothuis AM. Quality of general movements in infancy is related to neurological dysfunction, ADHD, and aggressive behaviour. Dev Med Child Neurol. 1999;41(6):381-91.
42. Fjortoft T, Einspieler C, Adde L, Strand LI. Inter-observer reliability of the "Assessment of Motor Repertoire--3 to 5 Months" based on video recordings of infants. Early Hum Dev. 2009;85(5):297-302.
43. Saether R, Stoen R, Vik T, Fjortoft T, Vagen RT, Silberg IE, et al. A change in temporal organization of fidgety movements during the fidgety movement period is common among high risk infants. Eur J Paediatr Neurol. 2016;20(4):512-7.
44. Cioni G, Bos AF, Einspieler C, Ferrari F, Martijn A, Paolicelli PB, et al. Early neurological signs in preterm infants with unilateral intraparenchymal echodensity. Neuropediatrics. 2000;31(5):240-51.
45. Guzzetta A, Mercuri E, Rapisardi G, Ferrari F, Roversi MF, Cowan F, et al. General movements detect early signs of hemiplegia in term infants with neonatal cerebral infarction. Neuropediatrics. 2003;34(2):61-6.
46. Bouwstra H, Dijk-Stigter GR, Grooten HM, Janssen-Plas FE, Koopmans AJ, Mulder CD, et al. Predictive value of definitely abnormal general movements in the general population. Dev Med Child Neurol. 2010;52(5):456-61.
47. Peyton C, Yang E, Msall ME, Adde L, Stoen R, Fjortoft T, et al. White Matter Injury and General Movements in High-Risk Preterm Infants. AJNR Am J Neuroradiol. 2017;38(1):162-9.
48. Butcher PR, van Braeckel K, Bouma A, Einspieler C, Stremmelaar EF, Bos AF. The quality of preterm infants' spontaneous movements: an early indicator of intelligence and behaviour at school age. J Child Psychol Psychiatry. 2009;50(8):920-30.
49. Heineman KR, Schendelaar P, Van den Heuvel ER, Hadders-Algra M. Motor development in infancy is related to cognitive function at 4 years of age. Dev Med Child Neurol. 2018;60(11):1149-55.
50. Einspieler C, Bos AF, Libertus ME, Marschik PB. The General Movement Assessment Helps Us to Identify Preterm Infants at Risk for Cognitive Dysfunction. Front Psychol. 2016;7:406.
51. Rahmati H, Aamo OM, Stavdahl O, Dragon R, Adde L. Video-based early cerebral palsy prediction using motion segmentation. Conf Proc IEEE Eng Med Biol Soc. 2014;2014:3779-83.
52. Rahmati H, Martens H, Aamo OM, Stavdahl O, Stoen R, Adde L. Frequency Analysis and Feature Reduction Method for Prediction of Cerebral Palsy in Young Infants. IEEE Trans Neural Syst Rehabil Eng. 2016;24(11):1225-34.
53. Gao Y, Long Y, Guan Y, Basu A, Baggaley J, Plötz T. Towards Reliable, Automated General Movement Assessment for Perinatal Stroke Screening in Infants UsingWearable Accelerometers. ACM. 2019. arXiv:1902.08068 [cs.HC]
54. Karch D, Kang KS, Wochner K, Philippi H, Hadders-Algra M, Pietz J, et al. Kinematic assessment of stereotypy in spontaneous movements in infants. Gait Posture. 2012;36(2):307-11.
55. Aristidou A, Cohen-Or D, Hodgins JK, A S. Self-similarity Analysis for Motion Capture Cleaning. Computer Grafics-Forum. 2018;37(2):297-309.
56. SCPE. Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. Surveillance of Cerebral Palsy in Europe (SCPE). Dev Med Child Neurol. 2000;42(12):816-24.

Figure 1: Setup

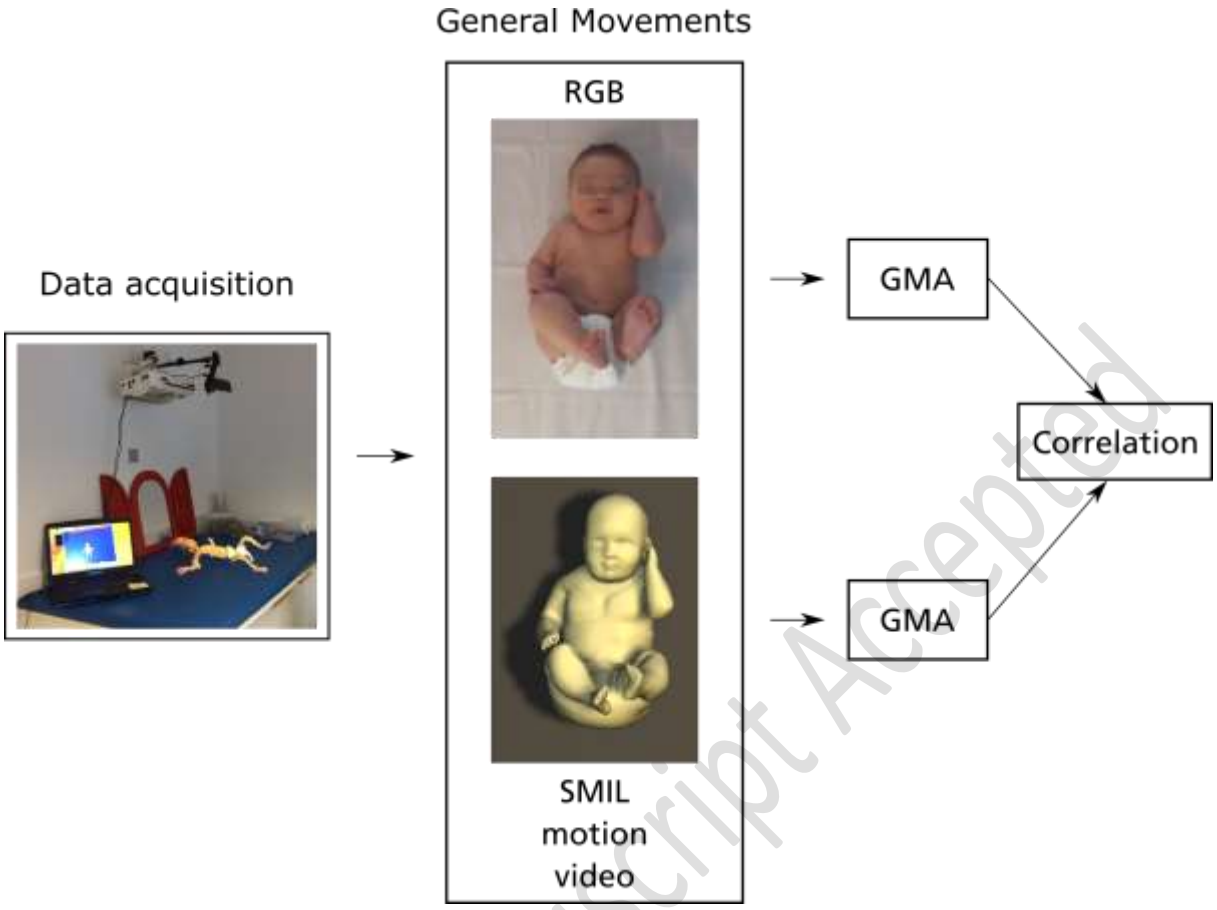
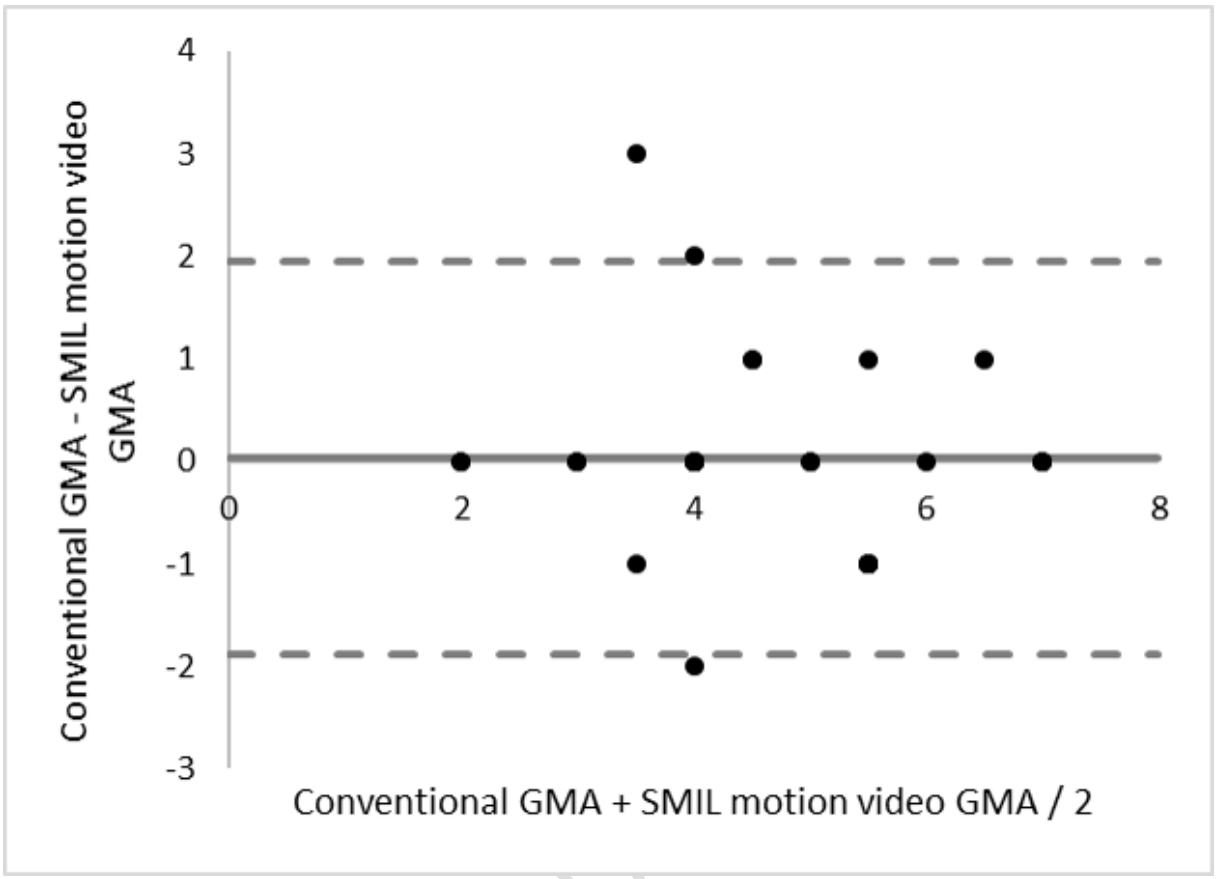
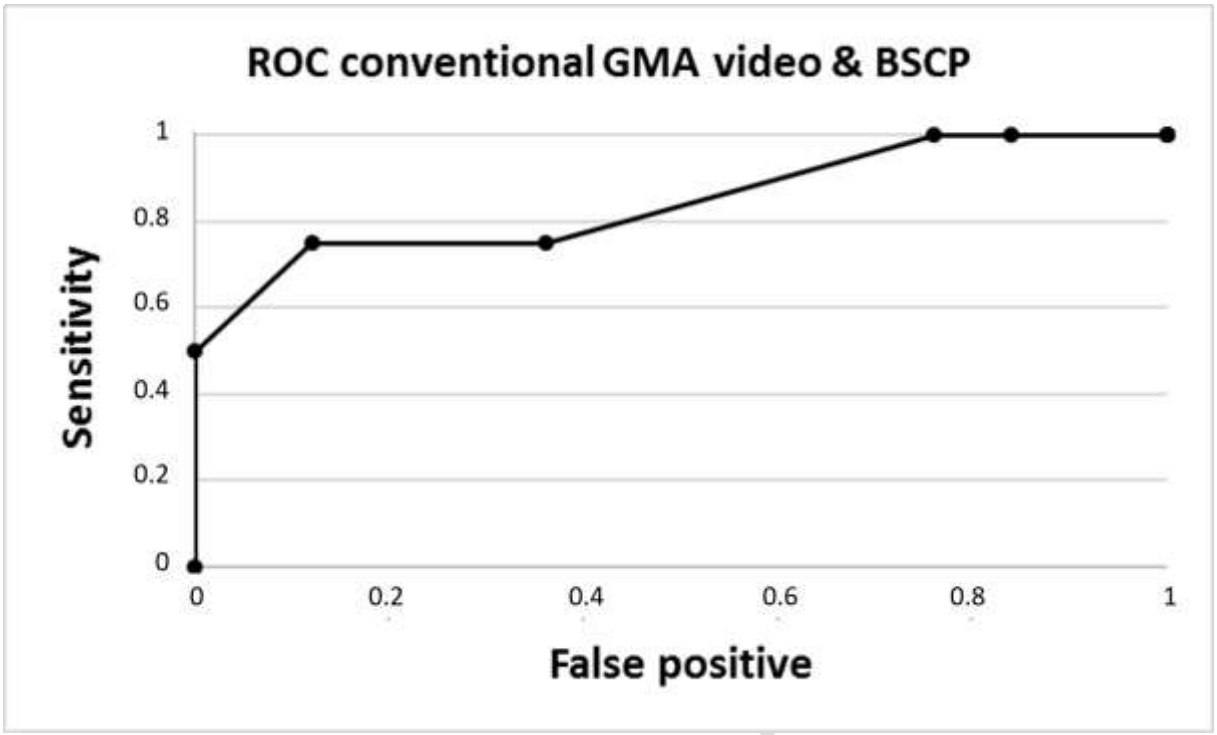


Figure 2: Conventional video recording and its corresponding SMIL motion video pose.

Supplementary Figure S1: Bland-Altman Plot: 1-10 Likert scale rating conventional GMA versus SMIL motion video rating.



Supplementary Figure S2a,b: ROC



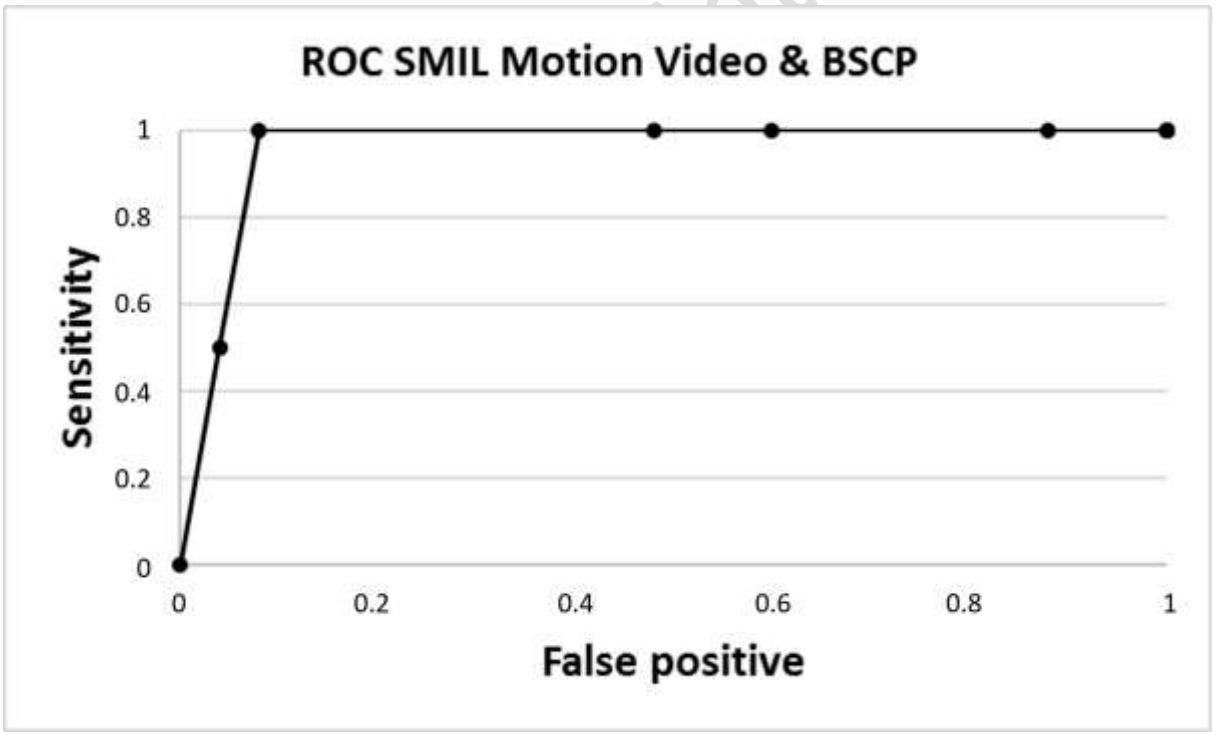


Table 1: GMA Conventional video *versus* SMIL motion video

Pat ID	sex	GA	CA	Clinical Diagnosis at 12-30 mo CA [38]	GM-complexity Likert score Conventional video	Fidgety movements Conventional video	GM-complexity Likert score SMIL motion video	Fidgety movements SMIL motion video
1	M	29+4	12	**Bilateral CP**	**2**	**0**	**2**	**0**
2	M	40+0	17	**Bilateral CP**	**2**	**0**	**2**	**1**
3	F	28+4	12	Dev. Delay	5	**2**	**2**	**1**
4	M	29+1	18	Dev. Delay	**3**	**2**	**3**	**2**
5	M	40+2	9	**Bilateral CP**	**3**	**1**	**3**	**2**
6	F	32+2	14	**Bilateral CP**	**5**	**0**	**3**	**0**
7	F	24+5	12	Dev. Delay	**3**	2	4	2
8	M	24+0	13	Dev. Delay	4	2	4	2
9	M	35+2	18	Age adequate	4	2	4	2
10	F	28+0	15	Dev. Delay	4	2	4	2
11	M	40+6	17	Age adequate	4	2	4	2
12	M	25+5	14	Dev. Delay	4	2	4	2
13	F	28+5	16	Age adequate	4	2	4	2
14	F	28+0	15	Dev. Delay	5	2	4	2
15	M	41+5	9	Age adequate	**3**	1	5	1
16	M	27+1	14	Dev. Delay	5	2	4	2
17	M	31+1	17	Age adequate	5	2	5	2
18	M	29+1	13	Age adequate	5	2	5	2
19	F	33+4	12	Age adequate	5	2	6	2
20	F	28+0	15	**Unilateral CP**	**5**	2	**6**	2
21	F	28+5	16	Age adequate	5	2	6	2
22	M	29+1	13	Age adequate	5	2	6	2
23	M	23+6	15	Dev. Delay	5	2	6	2
24	F	29+4	16	Dev. Delay	6	2	5	2
25	M	29+4	16	Dev. Delay	6	2	6	2
26	F	36+3	12	Dev. Delay	7	2	6	2
27	F	26+6	17	Age adequate	7	2	7	2
28	M	25+4	15	Dev. Delay	7	2	7	2
29	M	41+1	16	Dev. Delay	7	2	7	2

Legend: GM-complexity Likert scale ratings corrrespond to the following classifications 1-3 “definitely abnormal”, 4-5 “mildly abnormal”, 6-7 “normal suboptimal”, 8-10 “ normal optimal. Fidgety movements (FMs): 0 = absent; 1 = sporadically present; 2 = intermittently or continuously present. Abbreviations: CA = corrected age; convent = conventional; CP = cerebral palsy; definition of CP in accordance with international expert reccomendations.[38]; terminology of the Surveillance of Cerebral Palsy in Europe (SCPE; [56]. DD: developmental delay, either based on performance on the Bayley Scales of Infant and Toddler Development (BSITD, Version 3) or on the clinical examination (see Supplementary Table 1); GA: gestational age; F: female; M: male;

Table 2: Prediction of CP at follow up by conventional video GMA rating and GMA rating based on SMIL motion videos.

a) based on GM-complexity

		Cerebral palsy	No CP	Total	Q criteria
Conventional video					Sensitivity 0.6
DA-compl.		3	3	**6**	Specificity 0.875
Not-DA		2*	21	**23**	PPV 0.5
	Total	**5**	**24**	**29**	NPV 0.913
SMIL motion video					Sensitivity 0.8
DA compl.		4	2	**6**	Specificity 0.917
Not-DA		1*	22	**23**	PPV 0.667
	Total	**5**	**24**	**29**	NPV 0.957

b) based on FMs

		Cerebral palsy	No CP	Total	Q criteria
Conventional video					Sensitivity 0.6
FMs absent		3	0	**3**	Specificity 1
FMs present		2*	24	**26**	PPV 1
	Total	**5**	**24**	**29**	NPV 0.923
SMIL motion video					Sensitivity 0.4
FMs absent		2	0	**2**	Specificity 1
FMs present		3*	24	**27**	PPV 1
	Total	**5**	**24**	**29**	NPV 0.889

c) based on GM-complexity and FMs

		Cerebral palsy	No CP	Total	Q criteria
Conventional video					
DA-compl., FMs absent		2	0	**2**	Sensitivity 0.4
Not-DA, FMs absent		1	0	**1**	Specificity 1
DA-compl., FMs present		1	3	**4**	PPV 1
Not-DA, FMs present		1*	21	**22**	NPV 0.889
	Total	**5**	**24**	**29**	
SMIL motion video					
DA-compl, FMs absent		2	0	**2**	Sensitivity 0.4
Not-DA, FMs absent		0	0	**0**	Specificity 1
DA-compl, FMs present		2	2	**4**	PPV 1
Not-DA, FMs present		1*	22	**23**	NPV 0.889
	Total	**5**	**24**	**29**	

Legend Tabel 2a-c: * including infant #20 later diagnosed with unilateral spastic CP. Abbreviations: DA compl. = definitely abnormal complexity GM rating (Likert Scale 1-3); Not-DA = all other GM ratings (mildly abnormal, normal suboptimal; Likert scale 4-7). FMs absent = fidgety movments scored "0", FMs present = all other fidgety movements (sporadic, intermittend, continuously). [36, 37], SMIL Skinned Multi-Infant Linear Body Model [34].

Supplementary Table S1: Clinical characteristics of the study population

Pat ID	sex	GA	CA	BW	APGAR 1/5/10 minutes	Umbilical cord pH	Umbilical Base-excess	Cerebral ultrasound	Clinical Diagnosis at 12-30 months
1	M	29+4	12	1370	8/9/9	7.41	-2	Abnormal (PVL)	**Bilateral CP**
2	M	40+0	17	3500	1/2/6	7.14	-12	Abnormal (HIE term asphyxia)	**Bilateral CP**
3	F	28+4	12	800	8/9/9	7.3	-8	Unspecific	Bayley III: DD at 23 mo CA
4	M	29+1	18	960	2/8/9	7.36	1	Normal	Clinically DD at 18 mo CA
5	M	40+2	9	2500	6/7/7	7.08	-11	Abnormal (HIE term asphyxia)	**Bilateral CP**
6	F	32+2	14	1810	6/8/-	7.36	0	Abnormal (PVL)	**Bilateral CP**
7	F	24+5	12	820	7/9/9	7.42	2	Normal	Bayley III: DD at 24 mo CA
8	M	24+0	13	620	7/8/9	7.39	1	Unspecific	Clinically DD at 12 mo CA
9	M	35+2	18	2590	9/9/10	7.4	-2	Abnormal (ICH I right, II° left)	Typical development at 12 mo CA
10	F	28+0	15	480	7/8/10	7.19	-5.5	Normal	Bayley III: DD at 26 mo CA
11	M	40+6	17	4526	-/9/10	7.32	0	Unspecific	Typical development at 12 mo CA
12	M	25+5	14	750	1/2/4	7.31	-7	Unspecific	Clinically DD at12 mo CA
13	F	28+5	16	1300	9/9/10	7.47	-4.5	Unspecific	Bayley III: typical development at 28 mo CA
14	F	28+0	15	650	8/10/10	7.38	-6	Unspecific	Bayley III: DD at 24 mo CA
15	M	41+5	9	3450	3/6/10	7.23	-8.4	Normal	Bayley III: typical development at 24 mo CA
16	M	27+1	14	1100	4/9/9	7.31	-4.3	Normal	Bayley III: DDat 30 mo CA

17	M	31+1	17	-	8/7/9	7.33	-1	Unspecific	Bayley III: typical development at 26 mo CA
18	M	29+1	13	1190	8/8/9	7.34	-3.2	Normal	Bayley III: typical development at 24 mo CA
19	F	33+4	12	1850	9/10/10	7.47	-2.5	Normal	Typical development at 12 mo CA
20	F	28+0	15	900	9/9/9	7.35	-5.5	Abnormal (PVC left)	**Unilateral CP at 6 mo CA**
21	F	28+5	16	1170	5/8/9	7.08	-10.3	Normal	Bayley III: typical development at 28 mo CA
22	M	29+1	13	1320	7/8/9	7.36	-3	Unspecific	Bayley III: typical developmentat 24 mo CA
23	M	23+6	15	-	-	-	-	Unspecific	Bayley III: DD at 31 mo CA
24	F	29+4	16	1756	5/6/7	7.36	-4.9	Normal	Clinically DD at 12 mo CA
25	M	29+4	16	850	2/2/2	0	0	Normal	Clinically DD at 12 mo CA
26	F	36+3	12	840	1/5/8	7.4	0	Normal	Clinically DD 12 mo CA
27	F	26+6	17	890	6/8/10	7.42	-0.7	Unspecific	Bayley III: typical development at 25 mo CA
28	M	25+4	15	655	4/9/9	7.4	-2.1	Normal	Clinically DD at 12 mo CA
29	M	41+1	16	4280	1/3/5	6.79	-27.2	Unspecific	Clinically DD at 12 mo CA

Legend: Abbreviations: sex = male (m), female (f), GA = gestational age, CA = corrected age; BW = bodyweight, HIE = Hypoxic ischaemic encephalopathy, PVL = periventricular leucomalasia, PVC = periventricular cyst, CP = cerebral palsy, DD = developmental delay, either based on performance on the Bayley Scales of Infant and Toddler Development (BSITD, Version 3) or on the clinical examination mo = months. Definition of CP in accordance with international expert reccomendations.[38]; terminology of the Surveillance of Cerebral Palsy in Europe (SCPE; [56].

Legends to the figures

Figure 1 Study set-up

We recorded General Movements of Infants in the clinical setting of an outpatient clinic. We correlated expert GMA ratings of standard RGB videos with GMA ratings on SMIL motion videos of the same sequence. SMIL motion videos are the result of capturing 3D shape and motion with the SMIL model [34]. (further explanation see: https://www.youtube.com/watch?v=aahF1xGurmM&feature=youtu.be)

Figure 2: Conventional video recording and its corresponding SMIL motion video pose.

The upper panel shows 8 representative poses of a three minute video recording of an infant at fidgety age. The infant presents with typical motion features of complexity and variation during active wakefulness; his GMA complexity was rated as normal suboptimal (Likert-score 7). The lower panel are the corresponding SMIL motion video frames; also in these panels GM-complexity can be clearly observed. Note that fidgety movements can not be shown in the static poses of the figure.

Supplementary Figure S1: Bland-Altman Plot of correspondence of Likert-scale rating of GM-complexity comparing conventional RGB video and SMIL motion video.

The Bland-Altman plot depicts the agreement between the ratings of both videos. The Y axis shows the difference between the two paired measurements, the X-axis represents the average of these measures. Mean value of the 29 GMA rating differences = 0.034 (grey solid line). Standard deviation of mean value: ±0.981). Upper dotted line = upper reference limit (Mean value + 1.96 x Stddev.)= 1.958. Lower dotted line = lower reference limit (Mean value – 1.96 x Stddev) = -1.888. Regarding the three outliers infant #3 (developmental delay) infant #6 (bilateral spastic CP) and infant #15 refer to the discussion section.

Supplementary Figure S2: ROC of sensitivity of GMA DA complexity rating based on the two types of video

Fig S2a: Receiver operating characteristic curve (ROC curve) of DA vs non-DA complexity rating based on conventional RGB video; S2b: similar curve but now based on rating of SMIL motion video. The lower panel suggests a superior diagnostic ability of the SMIL motion video GM-complexity rating of the binary classification system.